

A STOCHASTIC MODEL FOR DIRECTED GRAPHS WITH
TRANSITION RATES DETERMINED BY RECIPROCITY

by

Stanley S. Wasserman*
Department of Applied Statistics
School of Statistics
University of Minnesota
Saint Paul, MN 55108

Technical Report No. 305

November, 1977

*Partial support provided by the National Science Foundation,
Grant No. SOC73-05489 to Carnegie-Mellon University.

A STOCHASTIC MODEL FOR
DIRECTED GRAPHS WITH
TRANSITION RATES DETERMINED BY
RECIPROCITY

Stanley S. Wasserman *
Department of Applied Statistics
University of Minnesota
St. Paul, Minnesota 55108

October 1977

First draft; comments are welcome.

* Partial support provided by National Science Foundation Grant
S0C73-05489 to Carnegie-Mellon University.

A STOCHASTIC MODEL FOR
DIRECTED GRAPHS WITH
TRANSITION RATES DETERMINED BY
RECIPROCITY

ABSTRACT

We present a stochastic model for the evolution of directed graphs over time. The infinitesimal transition rates for each arc in a directed graph depend only on the presence or absence of a reciprocated arc. The model reduces the entire directed graph to a set of $\binom{g}{2}$ independent and identically distributed dyad processes. The properties of the dyad process are discussed, four parameters of the stochastic model examined, and estimation strategies given based on maximum likelihood and the embeddability of the data for several sampling schemes.

KEYWORDS: Directed Graphs; Stochastic Modelling; Embeddability; Statistical Inference in Markov Chains.

This paper addresses the problem of mathematically describing the evolution of a binary directed graph over time. We discuss a simple stochastic model for this process of change, and describe several strategies for estimating the parameters of the model. The appropriate strategy for a given situation depends on the number of times the graph is observed in its evolution. The model has been successfully applied to several data sets by Wasserman (1977a).

The use of binary directed graphs as structural representations for processes is widespread. Wasserman (1977b) comments on a few areas of application, including sociology, communication, transportation, and several fields in the natural sciences. Most of these processes are acknowledged to change over time, an evolution that alters the arcs that exist between nodes. However, there has been little effort directed at the construction of realistic models for this evolution. Undirected graphs and multivalued graphs have been effectively modelled stochastically, but the fitting of these models to binary directed graphs yields little insight into the binary processes under investigation. Indeed, the most important features of these processes are the on/off nature and the directedness of the arcs in the structural graph. Ignoring these two qualities in the model construction would be a fundamental error.

The stochastic model described here is quite simple. But one should not assess the merits of such a model by its complexity or lack thereof. Much can be learned from simple, non-elaborate models, particularly in the social sciences. The information to be gained yields valuable insights into the process in question. Hopefully, such research will eventually allow researchers to postulate and adequately analyze more complicated models. As noted by Wasserman (1977a), this has certainly been true of the research discussed in this paper.

1. Introduction to the Problem

Consider a directed graph or digraph that structurally represents the state of some process at time t . A digraph is a set $V = \{v_1, v_2, \dots, v_g\}$ of nodes and a set $L = \{l_1, l_2, \dots, l_c\}$ of directed arcs connecting pairs of nodes. We let $l_i = v_j v_k$ be the directed line running from node v_j to node v_k and further stipulate that the digraph be binary; i.e., if two distinct arcs exist such that $l_i = v_j v_k$ and $l_{i'} = v_j v_k$, then $l_i = l_{i'}$. In addition, arcs $l_i = v_j v_j$ do not exist, ruling out the existence of loops in the digraph. We commonly let D_g represent a digraph with g nodes.

Let $\underline{X}(t)$ be the adjacency matrix representing the state of D_g at time t . Specifically, $\underline{X}(t) = [X_{ij}(t)]$, where

$$X_{ij}(t) = \begin{cases} 1 & \text{if } v_i v_j \in L \text{ at time } t, \\ 0 & \text{otherwise.} \end{cases}$$

The time parameter t is assumed to be continuous, $t \geq 0$.

The matrices $\underline{x}, \underline{w}, \underline{y}, \underline{z}, \dots$ are single states of the continuous time stochastic process $\underline{X}(t)$. The process has a state space \underline{S} of all possible $(g \times g)$ binary-valued matrices with zero diagonal -- $2^{g(g-1)}$ in number, making \underline{S} quite large.

The problem that this paper considers is the formulation and evaluation of a stochastic model for $\underline{X}(t)$ where the transitions between states of \underline{S} depend on the state of the digraph at time t . We accomplish this by assuming $\underline{X}(t)$ to be a continuous time Markov chain and by allowing the infinitesimal transition rates to be specific functions of the elements of $\underline{X}(t)$. This solution to the problem was first suggested by Holland and Leinhardt (1977a) and has been further elaborated on by Holland and Leinhardt (1977b) and by Wasserman (1977c). Several models using these assumptions are discussed in depth by Wasserman

(1977a), including the model incorporating only reciprocity which is the subject of this paper.

In section 2 we discuss the Holland-Leinhardt modelling framework, briefly outlining the assumptions of the framework. We present a simple model for digraphs based on reciprocity and describe how this particular parameterization substantially reduces the size of the state space. We compute moments and equilibrium distribution for the stochastic process arising from the reciprocity model in section 3.

The particular parameterization chosen for the reciprocity model has four parameters that can vary from digraph to digraph. In the latter sections of this paper, we estimate these parameters. In section 4, we consider whether an observed set of observations of a digraph is representable as a continuous time Markov chain, and if so, give a simple procedure for computing reasonable parameter estimates. Maximum likelihood estimation, a computationally more difficult procedure, is discussed in section 5.

2. The Modelling Framework

The Holland-Leinhardt framework consists of two assumptions regarding the stochastic nature of the arcs $X_{ij}(t)$. The first is that $\underline{X}(t)$ is a Markov chain:

Assumption 1. Markov

$\underline{X}(t)$ is a standard Markov chain with finite state space \underline{S} and

$$(1) \quad P_{\underline{xy}}(t, h) = P\{\underline{X}(t+h) = \underline{y} \mid \underline{X}(t) = \underline{x}\}$$

as the probability transition matrix.

Secondly, we assume that for small intervals of time $(t, t+h)$, the changes in the arcs of a digraph are statistically independent:

Assumption 2. Conditional Change Independence

$$(2) \quad P_{\underline{xy}}(t, h) = \prod_{i,j} P\{X_{ij}(t+h) = y_{ij} \mid \underline{X}(t) = \underline{x}\} + o(h)$$

as $h \rightarrow 0$.

Assumption 2 is crucial. It implies that the probability of any two arcs changing simultaneously is essentially zero. In a small interval of time, there are only two changes that can occur for a single arc: arcs present at time t may disappear at time $t + h$, and vice versa. Thus, we represent the probability of arc changes as

$$(3) \quad P\{X_{ij}(t+h) = 1 - x_{ij} \mid \underline{X}(t) = \underline{x}\} = h \lambda_{ij}(\underline{x}, t) + o(h)$$

as $h \rightarrow 0$.

Note that λ , the change rate function, depends on the state of the digraph at time t , and on t itself.

Let $q_{\underline{xy}}(t)$ be the infinitesimal transition rates of $\underline{X}(t)$, the digraph process. We have

$$q_{\underline{xy}}(t) = \begin{cases} \lambda_{ij}(\underline{x}, t) & \text{if } \underline{y} \text{ and } \underline{x} \text{ differ only in the} \\ & (i, j)\text{th element} \\ 0, & \text{if } \underline{y} \text{ and } \underline{x} \text{ differ by more than one} \\ & \text{element} \end{cases}$$

(4)

and

$$q_{\underline{xx}}(t) = - \sum_{\underline{y} \neq \underline{x}} q_{\underline{xy}}(t) = - \sum_{i,j} \lambda_{ij}(\underline{x}, t)$$

as terms of $Q(t)$, the matrix of infinitesimal transition rates. Wasserman (1977a) discusses some characteristics of \underline{S} , and proves that if a simple condition is satisfied, specifically a restriction to non-zero change rates $\lambda_{ij}(\underline{x}, t)$, then the digraph process has an equilibrium distribution.

3. The Reciprocity Model and the Dyad Process

One can postulate various functional forms for the change rate function λ , defined in equation (3), and generate many models, all incorporating the two assumptions (1) and (2) of the modelling framework. We propose one such model in this section with a parameterization that produces a state space \underline{D} of only four states. The model is for reciprocity where the tendency over time for the arc $v_i v_j$ to exist depends only on the presence or absence of arc $v_j v_i$. We consider the $\binom{g}{2}$ dyads, or 2-subgraphs, of D_g , which by the model are independent and identically distributed. We compute moments and equilibrium distribution of the dyad process, and briefly describe its probability transition matrix.

We now dichotomize the change rate function (3) into two functions to allow for both types of arc changes. Define

$$(5a) \quad \lambda_{0ij}(\underline{x}) = \frac{1}{h} P\{X_{ij}(t+h) = 1 \mid \underline{X}(t) = \underline{x}, X_{ij}(t) = 0\} + o(1)$$

and

$$(5b) \quad \lambda_{1ij}(\underline{x}) = \frac{1}{h} P\{X_{ij}(t+h) = 0 \mid \underline{X}(t) = \underline{x}, X_{ij}(t) = 1\} + o(1)$$

for small h . Note that the rates of the process are now time homogeneous, and consequently, the digraph process stationary in time.

For the reciprocity model, we assume that

$$(6a) \quad \lambda_{0ij}(\underline{x}) = \lambda_0 + \mu_0 x_{ji}$$

and

$$(6b) \quad \lambda_{1ij}(\underline{x}) = \lambda_1 + \mu_1 x_{ji}.$$

There are $g(g-1)$ pairs of the change intensities (6a) and (6b) such that the pair for (i,j) depends only on the pair for (j,i) . The parameters λ_0 and λ_1 are measures of the overall rate of change for an arc, and μ_0 and μ_1 measure

the "importance" of a reciprocated arc. We expect that

$$(7a) \quad \lambda_0 > \lambda_1 > 0$$

and

$$(7b) \quad \begin{aligned} \mu_0 &> 0 \\ -\lambda_1 &< \mu_1 < 0 \end{aligned}$$

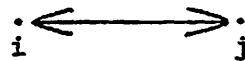
as argued from a consideration of various applications in Wasserman (1977a).

Let

$$(8) \quad D_{ij}(t) = [X_{ij}(t), X_{ji}(t)], i > j,$$

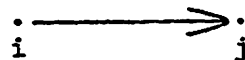
be the dyad for the pair of nodes (i,j). The state space \underline{D} of the dyad $D_{ij}(t)$ contains four states as illustrated in Figure 1. The Q intensity matrix for the dyad process is shown in Table 1.

$$D_{ij}(t) = (1,1)$$



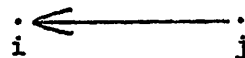
Mutual Relation (M)

$$D_{ij}(t) = (1,0)$$



Asymmetric Relation (A1)

$$D_{ij}(t) = (0,1)$$



Asymmetric Relation (A2)

$$D_{ij}(t) = (0,0)$$



Null Relation (N)

FIGURE 1. States Contained in \underline{D}

		S T A T E			
S		(0,0)	(1,0)	(0,1)	(1,1)
T	(0,0)	$-2\lambda_0$	λ_0	λ_0	0
A	(1,0)	λ_1	$-(\lambda_0 + \lambda_1 + \mu_0)$	0	$\lambda_0 + \mu_0$
T	(0,1)	λ_1	0	$-(\lambda_0 + \lambda_1 + \mu_0)$	$\lambda_0 + \mu_0$
E	(1,1)	0	$\lambda_1 + \mu_1$	$\lambda_1 + \mu_1$	$-2(\lambda_1 + \mu_1)$

TABLE 1. Q Matrix for the Dyad Process

The parameterization (6a) and (6b) of the reciprocity model yields a set of dyad processes $\{D_{ij}(t)\}$ that are independent. The entire $\underline{X}(t)$ digraph process can be represented as $\binom{g}{2}$ independent dyad processes consisting of the symmetrically positioned pairs of off-diagonal elements of $\underline{X}(t)$. Moreover the $\{D_{ij}(t)\}$ are identically distributed, continuous time, 4 state Markov chains, with state space \underline{D} .

We now consider the moments of the dyad process. Let

$$(9) \quad \underline{m}(t) = [m_{ij}(t), m_{ji}(t), m_{ij;ji}(t)]' \\ = [E\{X_{ij}(t)\}, E\{X_{ji}(t)\}, E\{X_{ij}(t)X_{ji}(t)\}]'$$

be the vector of first and second moments of the $\{D_{ij}(t)\}$. Also define

$$(10) \quad \underline{A} = \begin{bmatrix} -(\lambda_0 + \lambda_1) & \mu_0 & -(\mu_0 + \mu_1) \\ \mu_0 & -(\lambda_0 + \lambda_1) & -(\mu_0 + \mu_1) \\ \lambda_0 + \mu_0 & \lambda_0 + \mu_0 & -2(\lambda_0 + \mu_0 + \lambda_1 + \mu_1) \end{bmatrix} \\ \underline{b} = [\lambda_0, \lambda_0, 0]'$$

The following theorem gives a differential equation for $\underline{m}(t)$ and its solution.

Theorem 1. If the equations (6a) and (6b) are assumed as functions for the change rates (3), then the moments $\underline{m}(t)$ satisfy the equation

$$\frac{d\underline{m}(t)}{dt} = \underline{A} \underline{m}(t) + \underline{b}$$

which has solution

$$(11) \quad \underline{m}(t) = \underline{n}(t) + \int_{t_0}^t e^{(t-s)\underline{A}} \underline{b} ds$$

subject to the condition

$$\underline{m}(t_0) = \underline{n}(t_0) = [m_{01}, m_{02}, m_{03}]'$$

where

$$\frac{d\underline{n}(t)}{dt} = \underline{A} \underline{n}(t).$$

Proof. See Wasserman (1977a), as the necessary calculations are too lengthy to be reproduced here.

Calculation of the integral in equation (11) yields a very complicated solution for $\underline{m}(t)$ that gives little insight into the process. However, the equilibrium distribution of the dyad process is simple to calculate and easy to comprehend. After some necessary definitions, we show that the dyad process is reversible, allowing the equilibrium probabilities to be simply found from the reversibility equations.

Let $(q\{D_{ij}(t); D_{ij}(t+h)\})$ be the elements of the Q matrix for the dyad process given in Table 1. We define

$$(12) \quad \begin{aligned} \pi_M(t) &= P\{D_{ij}(t) = (1,1)\} \quad (\text{Mutual}) \\ \pi_{A1}(t) &= P\{D_{ij}(t) = (1,0)\} \quad (\text{Asymmetric}) \\ \pi_{A2}(t) &= P\{D_{ij}(t) = (0,1)\} \quad (\text{Asymmetric}) \\ \pi_N(t) &= P\{D_{ij}(t) = (0,0)\} \quad (\text{Null}) \end{aligned}$$

as the elements of $\underline{\pi}(t)$, the vector of probabilities of the 4 states of the dyad process. To prove the reversibility of $D_{ij}(t)$, we need the following

lemma that states the Kolmogorov cycle condition [Kendall (1959), equation 5.12] is a necessary and sufficient condition for reversibility.

Lemma 1. A necessary and sufficient condition for the continuous time Markov chain $D_{ij}(t)$ to be reversible is that, for any distinct

$$(13) \quad \begin{aligned} &\text{dyads } D_{ij}^{(1)}, D_{ij}^{(2)}, \dots, D_{ij}^{(n)}, n \leq 4, \text{ in } \underline{D}, \\ &q\{D_{ij}^{(1)}; D_{ij}^{(2)}\}q\{D_{ij}^{(2)}; D_{ij}^{(3)}\} \dots q\{D_{ij}^{(n)}; D_{ij}^{(1)}\} = \\ &q\{D_{ij}^{(1)}; D_{ij}^{(n)}\}q\{D_{ij}^{(n)}; D_{ij}^{(n-1)}\} \dots q\{D_{ij}^{(2)}; D_{ij}^{(1)}\} \end{aligned}$$

Proof: See Reich (1957), Theorem 2.

We call the closed path $(D_{ij}^{(1)}, D_{ij}^{(2)}, \dots, D_{ij}^{(n)})$ a cycle. It is not necessary to check (13) for every possible cycle, since all cycles are made up of certain simple paths. In fact, there is only one simple path for the dyad process, and that is

$$(14) \quad (1,1) \rightarrow (1,0) \rightarrow (0,0) \rightarrow (0,1) \rightarrow (1,1).$$

The following theorem proves that the dyad process is reversible and gives the equilibrium probabilities of the process.

Theorem 2. The dyad process is reversible, and the equilibrium probabilities are

$$(15) \quad \begin{aligned} \pi_M^{(\infty)} &= \frac{\lambda_0(\lambda_0 + \mu_0)}{(\lambda_0 + \lambda_1)(\lambda_1 + \mu_1) + \lambda_0(\lambda_0 + \mu_0 + \lambda_1 + \mu_1)} \\ \pi_{A1}^{(\infty)} &= \pi_{A2}^{(\infty)} = \frac{\lambda_0(\lambda_1 + \mu_1)}{(\lambda_0 + \lambda_1)(\lambda_1 + \mu_1) + \lambda_0(\lambda_0 + \mu_0 + \lambda_1 + \mu_1)} \\ \pi_N^{(\infty)} &= \frac{\lambda_1(\lambda_1 + \mu_1)}{(\lambda_0 + \lambda_1)(\lambda_1 + \mu_1) + \lambda_0(\lambda_0 + \mu_0 + \lambda_1 + \mu_1)} \end{aligned}$$

Proof: We use Lemma 1 and check reversibility for the one simple Kolmogorov cycle (14). We must show that the two following products are equal:

$$(16) \quad q\{(1,1); (1,0)\}q\{(1,0); (0,0)\}q\{(0,0); (0,1)\}q\{(0,1); (1,1)\}$$

and

$$(17) \quad q\{(1,1); (0,1)\}q\{(0,1); (0,0)\}q\{(0,0); (1,0)\}q\{(1,0); (1,1)\}$$

Examination of the elements of Table 1 shows that, term by term, the two products (16) and (17) are equal.

Because the dyad process is reversible, the following reversibility equations [Kendall (1959)] hold:

$$(18) \quad \begin{aligned} \pi_N(\infty)q\{(0,0); (0,1)\} &= \pi_{A2}(\infty)q\{(0,1); (0,0)\} \\ \pi_N(\infty)q\{(0,0); (1,0)\} &= \pi_{A1}(\infty)q\{(1,0); (0,0)\} \\ \pi_{A1}(\infty)q\{(1,0); (1,1)\} &= \pi_M(\infty)q\{(1,1); (1,0)\} \\ \pi_{A2}(\infty)q\{(0,1); (1,1)\} &= \pi_M(\infty)q\{(1,1); (0,1)\} \end{aligned}$$

where $\pi_N(\infty) + \pi_{A2}(\infty) + \pi_{A1}(\infty) + \pi_M(\infty) = 1$. Solving these equations (18) for $\pi(\infty)$ yields (15). Q.E.D.

Alternatively, we could have computed the equilibrium probabilities (15) by letting $t \rightarrow \infty$ in expression (11). Or, we could have set the system of differential equations for $\underline{m}(t)$ in Theorem 1 to zero, and solved for $\underline{m}(\infty)$.

The probability transition matrix for the dyad process is a very complicated expression. We find it by examining the eigenvalues and eigenvectors of the (4×4) infinitesimal generator \underline{Q} , and compute $e^{t\underline{Q}}$. The reader is referred to Wasserman (1977a) for the details.

4. Embeddability

We now discuss how one determines whether the reciprocity model provides a good description of the evolution of a directed graph. We review the notion of embeddability, the determination of whether an observed probability transition matrix could have arisen from a continuous time Markov model. We continue this discussion in the next section where we discuss estimation of the model parameters and evaluation of the fit of the model to data.

Suppose that we have observations on some process $z(t)$ denoted by $z_1(t)$, $z_2(t)$, ..., $z_K(t)$. Assume that the process has finite state space \underline{Z} , with states labelled $1, 2, \dots, N$. For example, if $z(t) = D_{ij}(t)$, the dyad process, we have $K = \binom{g}{2}$, $\underline{Z} = \underline{D}$, and $N = 4$. These observations $\{z(t)\}$ are collected at times $t = t_0, t_1, \dots, t_n$ where the t_j are distinct positive numbers.

If n is finite and nonzero, we define the empirical probability transition matrix $\hat{P}(t_\ell - t_k)$, $0 \leq k < \ell \leq n$, with elements $(\hat{p}_{ij}(t_\ell - t_k))$ as follows (where $t = t_\ell - t_k$):

$$(19) \quad \hat{p}_{ij}(t) = \frac{T_{ij}(t)}{\sum_j T_{ij}(t)}$$

where

$$T_{ij}(t) = \text{number of } z' \text{'s in state } i \text{ at time } t_k \text{ and in state } j \text{ at time } t_\ell.$$

These estimates are maximum likelihood estimates of the elements of the probability transition matrix of a stationary discrete time Markov chain [Anderson and Goodman (1957)]. For moderate n , there is a substantial number of estimated transition matrices that can be used to test the suitability of a continuous time Markov chain for these data.

Embeddability was first posed as a problem by Elfving (1937) and discussed by Kingman (1962), but only recently has the problem been rigorously solved by Singer and Spilerman (1974). The embedding problem, as formulated by Singer and Spilerman (1976), is:

Find simple test criteria on the elements of an observed stochastic matrix $\hat{P}(t)$, $0 < t < \infty$, which will guarantee that it can be written in the form $\hat{P}(t) = e^{tQ}$ for some Q with

$$(20) \quad \begin{aligned} q_{ij} &\geq 0, \text{ for } i \neq j, \\ q_{ii} &\leq 0, \text{ for } i = 1, 2, \dots, N, \\ \sum_j q_{ij} &= 0, \text{ for } i = 1, 2, \dots, N. \end{aligned}$$

The data analytic problem is to find the subclass of all Q matrices with structure (20) that could have given rise to the observed \hat{P} matrix or matrices.

This subclass of Q matrices is

$$(21) \quad Q = \{ \text{all matrices } Q \text{ with structure (20) such that } (t_i - t_j)Q = \log \hat{P}(t_i - t_j) \text{ for } 0 \leq j < i \leq n \}$$

Q may contain more than one element even when $n = 1$ because the logarithm function of a matrix is a "one to many" function. Moreover, when $n > 1$, the collection of Q matrices compatible with a subset of the empirical \hat{P} matrices may not coincide with the Q matrices compatible with a different subset. This situation is caused by the negation of the Markov assumption of time-homogeneous transition rates.

There are many necessary conditions for Q to be nonempty. Unfortunately, no simple sufficiency criteria exist. One must verify that the observed \hat{P} matrix satisfies the necessary conditions, and even then, there may not be a Q matrix compatible with the observed transition matrix.

Fortunately, if $\hat{\underline{P}}$ has positive, distinct, real eigenvalues, then there is a unique \underline{Q} matrix for a given $\hat{\underline{P}}$ since the scalar logarithm function is multiple valued only when it has complex arguments, and of course undefined with negative arguments. The following theorem, due to Sylvester, and also given by Singer and Spilerman (1976), states the result for this special instance.

Theorem 3. Sylvester's Theorem: If $\hat{\underline{P}}$ is an $(N \times N)$ matrix with distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_N$, and if $f(\cdot)$ is single valued in a neighborhood of each of the eigenvalues, then

$$(22) \quad f(\hat{\underline{P}}) = \sum_{i=1}^N f(\lambda_i) \prod_{j \neq i} (\hat{\underline{P}} - \lambda_j \underline{I}) / (\lambda_i - \lambda_j).$$

Proof: Gantmacher (1960) or Sylvester (1883).

The application of this theory to the reciprocity model is straightforward. For example, suppose that we assume the model is operating, and that we have observations on the digraph at time t_0 and t_1 . We reduce $\underline{X}(t_0)$ and $\underline{X}(t_1)$ to two sets of $\begin{pmatrix} g \\ 2 \end{pmatrix}$ dyads, and form $\hat{\underline{P}}(t_1 - t_0)$, a 4×4 matrix of dyad transitions. If this empirical probability transition matrix has 4 positive, real, distinct eigenvalues, we can calculate the unique empirical $\hat{\underline{Q}}$ matrix for these observations. Since we know the functional form of \underline{Q} (Table 1) we can easily solve for the parameters $\lambda_0, \lambda_1, \mu_0, \mu_1$ by setting the theoretical elements of \underline{Q} equal to the observed, calculated values.

Wasserman (1977a) gives conditions on the elements of $\hat{\underline{P}}$ that insure that the eigenvalues of $\hat{\underline{P}}$ are real, distinct, and positive. The conditions generally hold if the diagonal elements of $\hat{\underline{P}}$ are large, relative to the off-diagonal

elements. This "diagonal-dominant" situation is likely to occur when the two observation time points t_0 and t_1 are close. Taking observations close in time allows only a few transitions to other states so that the diagonal elements remain near unity.

We should note that occasionally $\log \hat{\underline{P}}$ yields a $\hat{\underline{Q}}$ which does not have structure (20); i.e., the off-diagonal elements of $\hat{\underline{Q}}$ may be negative. If so, then one may force the negative elements to zero, adjust the remaining terms to maintain the zero row sum, and obtain approximate parameter estimates. Or, a different $\hat{\underline{P}}$ can be examined if there are more than 2 observations on the process. This strategy of estimating \underline{Q} by computing logarithms of empirical transition matrices is flexible, but yields good estimates. We discuss the usefulness of this technique in the next section.

5. Parameter Estimation

In this section we discuss estimation of the parameters of the reciprocity model. There are three possible sampling schemes:

- 1) Continuous record,
- 2) Single observation,
- 3) Two or more observations.

We discuss each of these situations in turn.

5.1 Continuous Record

When one has a continuous record of all changes in each of the $\binom{g}{2}$ dyads for a time interval (t_0, τ) , estimation is relatively easy and will not be discussed. Billingsley (1961) discusses the theory for general Markov processes. One assumes that the Q matrix is a function of a vector of parameters $\underline{\theta}$, and reduces the continuous record $\{D_{ij}(t), t_0 \leq t \leq \tau\}$ for all dyads to observations on the discrete jumps of the process and the total waiting time in each state. Billingsley estimates $\underline{\theta}$ by maximum likelihood, and discusses likelihood ratio tests.

5.2 Single Observation

When one has a finite number of observations on a continuous time Markov chain, estimation is rather difficult. Darwin (1956) recognizes this, and states that the complex form of the likelihood, with observations on the process taken once or at regular intervals, almost prohibits the use of maximum likelihood estimation. Keiding (1974, 1975) considers sampling of the process at equidistant time points $t_0, t_j, t_{2j}, \dots, t_{Nj}$,

but only derives asymptotic results for $N \rightarrow \infty$.

We now consider estimation of the four parameters of the reciprocity model, $\lambda_0, \lambda_1, \mu_0, \mu_1$, assuming that we have a single observation on the entire digraph. A single adjacency matrix is the data set most frequently collected by sociologists studying social networks. We discuss the likelihood function of the parameters, given this single adjacency matrix, and estimate two meaningful functions of the parameters.

Suppose we have a sole observation of the stationary $\underline{X}(t)$ process, which we label \underline{x} . The matrix \underline{x} contains $\binom{g}{2}$ independent observations of the dyad process $D_{ij}(t)$ with state space \underline{D} . We are interested in the numbers of each type of dyad. Since the labelling of the nodes in the directed graph is arbitrary, we cannot distinguish a (1,0) asymmetric from a (0,1) asymmetric. Consequently, the information in \underline{x} can be summarized by three statistics:

$$\begin{aligned}
 & 1) \quad M(t) = \sum_{i > j} x_{ij} x_{ji} = \text{number of mutuals}, \\
 (23) \quad & 2) \quad A(t) = \sum_{i > j} [(1-x_{ij})x_{ji} + x_{ij}(1-x_{ji})] = \text{number of asymmetrics}, \\
 & 3) \quad N(t) = \sum_{i > j} (1-x_{ij})(1-x_{ji}) = \text{number of nulls},
 \end{aligned}$$

where $M(t) + A(t) + N(t) = \binom{g}{2}$.

Define $L(\underline{\theta} | \underline{x})$ as the likelihood function of the parameters given the single matrix \underline{x} , where $\underline{\theta} = (\lambda_0, \lambda_1, \mu_0, \mu_1)'$. L is merely a multinomial likelihood:

$$(24) \quad L(\underline{\theta} | \underline{x}) = \prod_{\substack{\text{all dyads} \\ i > j}} \pi_M(t)^{x_{ij}x_{ji}} \pi_{A1}(t)^{x_{ij}(1-x_{ji})} \pi_{A2}(t)^{(1-x_{ij})x_{ji}} \pi_N(t)^{(1-x_{ij})(1-x_{ji})}$$

where the π 's, defined in equation (12), are the probabilities of the four states of the dyad process. The log likelihood can be written succinctly as:

$$(25) \quad \log L(\underline{\theta} | \underline{x}) = M(t) \log \pi_M(t) + A(t) \log \pi_A(t) + N(t) \log \pi_N(t)$$

where $\pi_A(t) = \pi_{A1}(t) + \pi_{A2}(t)$, and indeed, $M(t)$, $A(t)$, and $N(t)$ are sufficient statistics.

The expressions for π_M , π_A , and π_N are linear combinations of the elements of the $\underline{m}(t)$ vector of moments, given in (11), and hence, are quite complicated expressions. For computational ease, we shall use the steady state values of π_M , π_A , and π_N , defined in (15).

$$(26) \quad \begin{aligned} \text{Let } d &= \lambda_0(\lambda_0 + \lambda_1 + \mu_0 + \mu_1) + (\lambda_0 + \lambda_1)(\lambda_1 + \mu_1) \text{ so that} \\ \pi_M^{(\infty)} &= \frac{\lambda_0(\lambda_0 + \mu_0)}{d} = \frac{a}{a + 2b + c} \\ \pi_A^{(\infty)} &= \frac{2\lambda_0(\lambda_1 + \mu_1)}{d} = \frac{2b}{a + 2b + c} \\ \pi_N^{(\infty)} &= \frac{\lambda_1(\lambda_1 + \mu_1)}{d} = \frac{c}{a + 2b + c} \end{aligned}$$

The log likelihood (25) is:

$$(27) \quad \log L(\underline{\theta} | \underline{x}) = M(t) \log (a/2b) - N(t) \log (2b/c) - \left(\frac{g}{2} \right) \log (a/2b + c/2b + 1)$$

where $A(t) = \begin{pmatrix} g \\ 2 \end{pmatrix} - M(t) - N(t)$.

The likelihood function (27) depends on 4 unknown parameters $\underline{\theta}$, but contains only 2 "pieces of information": $M(t)$ and $N(t)$. Hence, we can only estimate 2 functions of the parameters, which are

$$(28) \quad \begin{aligned} \theta_1 &= a/b = (\lambda_0 + \mu_0) / (\lambda_1 + \mu_1) \\ \theta_2 &= b/c = \lambda_0 / \lambda_1 \end{aligned}$$

θ_1 is the ratio of the probabilities of change, in a small time interval, in the presence of a reciprocated arc, and θ_2 , the ratio in the absence of a reciprocated arc. The maximum likelihood estimates of θ_1 and θ_2 [see Wasserman (1977a)] are

$$(29) \quad \begin{aligned} \hat{\theta}_1 &= 2[M(t)/A(t)] \\ \hat{\theta}_2 &= 1/2[A(t)/N(t)] \end{aligned}$$

and are easily computed from the data.

There are two other ratios that are more interesting than θ_1 and θ_2 . These are

$$(30) \quad \begin{aligned} \kappa_1 &= \frac{\lambda_0 + \mu_0}{\lambda_0} \\ \kappa_2 &= \frac{\lambda_1 + \mu_1}{\lambda_1} \end{aligned}$$

κ_1 and κ_2 directly emphasize the importance of a reciprocated arc. κ_1 (κ_2) is the effect of a reciprocated arc on the change from a non-choice to a choice (choice to a non-choice). We suspect that $\kappa_1 > 1$ and $\kappa_2 < 1$. Unfortunately, these "reciprocity" ratios cannot be estimated via maximum

likelihood with only one observation on the digraph [see Wasserman (1977a) for proof of this fact].

Note that the ratio of the change ratios θ_1 and θ_2 is an odds ratio.

We have

$$\frac{\theta_1}{\theta_2} = \frac{\kappa_1}{\kappa_2} =$$

$$(31) \frac{P\{X_{ij}(t+h) = 1 | D_{ij}(t) = (0,1)\}}{P\{X_{ij}(t+h) = 0 | D_{ij}(t) = (1,1)\}} \bigg/ \frac{P\{X_{ij}(t+h) = 1 | D_{ij}(t) = (0,0)\}}{P\{X_{ij}(t+h) = 0 | D_{ij}(t) = (1,0)\}} =$$

$$\frac{P\{X_{ij}(t) = 0, X_{ij}(t+h) = 1 | X_{ji}(t) = 1\}}{P\{X_{ij}(t) = 1, X_{ij}(t+h) = 0 | X_{ji}(t) = 1\}} \bigg/ \frac{P\{X_{ij}(t) = 0, X_{ij}(t+h) = 1 | X_{ji}(t) = 0\}}{P\{X_{ij}(t) = 1, X_{ij}(t+h) = 0 | X_{ji}(t) = 0\}}$$

the increase in the odds of a new arc $v_i v_j$ coming into existence during the interval $(t, t+h)$ due to the existence of $v_j v_i$. With no reciprocity effect, $\log(\theta_1/\theta_2) = 0$; if $\log(\theta_1/\theta_2) > 0$, a positive reciprocity effect is present. So even though θ_1 and θ_2 considered separately are not very informative, their ratio is quite interesting.

5.3 Two or More Observations

We now examine situations where one has a data set containing several observations on the directed graph. First consider $n = 2$, and let $\underline{X}(t_1)$ and $\underline{X}(t_2)$, $t_2 > t_1$, be the two observations on the process. We examine each of the pairs $(X_{ij}(t_1), X_{ji}(t_1))$ and $(X_{ij}(t_2), X_{ji}(t_2))$ in turn, and form a 4×4 contingency table, with rows corresponding to the dyad state at time t_1 , and columns to the t_2 state. There are $\binom{g}{2}$ "counts" in this table [See Bishop, Fienberg, and Holland (1975) for discussion of representation of data from Markov models as contingency tables].

We let \underline{T} denote such a table, with entries $(t_{k\ell})$, where

$$\begin{aligned}
 k = 1 &= \text{Null } (0,0) \\
 k = 2 &= \text{Asymmetric } (1,0) \\
 k = 3 &= \text{Asymmetric } (0,1) \\
 k = 4 &= \text{Mutual } (1,1)
 \end{aligned}
 \tag{32}$$

and similarly for the subscript ℓ .

The likelihood function when $n = 2$ is

$$\begin{aligned}
 L(\theta | \underline{X}(t_1) = \underline{x}, \underline{X}(t_2) = \underline{y}) = \\
 [\pi_M(t_1)^{M(t_1)} \pi_A(t_1)^{A(t_1)} \pi_N(t_1)^{N(t_1)}] \prod_{k,\ell} p_{k\ell}(t_2 - t_1)^{t_{k\ell}}
 \end{aligned}
 \tag{33}$$

where the $(p_{k\ell}(t_2 - t_1))$ are the elements of the probability transition matrix for the dyad process.

Suppose $n > 2$. We now have observations on the digraph process $\underline{X}(t_1)$, $\underline{X}(t_2), \dots, \underline{X}(t_n)$. We form $(n-1)$ matrices \underline{T}_m , $m = 1, 2, \dots, (n-1)$, with elements $(t_{k\ell m})$, where \underline{T}_m gives the transitions of the dyads at time t_m to the dyads at time t_{m+1} . The process is stationary, so that the probability transition matrix for the dyads depends only on $t_{m+1} - t_m$. The likelihood function is

$$\begin{aligned}
 L(\theta | \underline{X}(t_1) = \underline{x}_1, \underline{X}(t_2) = \underline{x}_2, \dots, \underline{X}(t_n) = \underline{x}_n) = \\
 [\pi_M(t_1)^{M(t_1)} \pi_A(t_1)^{A(t_1)} \pi_N(t_1)^{N(t_1)}] \cdot \\
 \prod_{m=1}^{n-1} \prod_{k,\ell} p_{k\ell}(t_{m+1} - t_m)^{t_{k\ell m}}
 \end{aligned}
 \tag{34}$$

When we have equidistant sampling, such that $t_2 - t_1 = t_3 - t_2 = \dots = t_n - t_{n-1} = t$, then the likelihood (34) simplifies to

$$(35) \quad L(\underline{\theta} | \underline{x}_1, \underline{x}_2, \dots, \underline{x}_n) = \\ [\pi_M(t_1)^{M(t_1)} \pi_A(t_1)^{A(t_1)} \pi_N(t_1)^{N(t_1)}] \prod_{k,l} p_{kl}(t)^{\sum_m t_{klm}}.$$

We can "pool" transitions across time points if we are certain of time homogeneity, so that we have an effective sample of $(n-1) \binom{g}{2}$ dyads.

To estimate the four parameters $\underline{\theta}$, we can differentiate the logarithm of (33) or (34) with respect to each of the four parameters, setting the resulting derivatives to zero, to obtain a system of 4 equations in 4 unknowns. Unfortunately, the p_{kl} 's are highly nonlinear, being sums of exponentials, and maximum likelihood estimation in such a situation is not only unreliable, but the solutions can only be obtained approximately.

However, we may study the embeddability of the data in hand, and develop a new strategy. If $n = 2$, we compute the (4×4) empirical probability transition matrix and using the rules outlined in the previous section, find an empirical \underline{Q} matrix, $\hat{\underline{Q}}$. The elements of \underline{Q} are simple functions of the parameters $\lambda_0, \lambda_1, \mu_0, \mu_1$; consequently, reliable estimates of the parameters are easily obtained. If $n > 2$, we have several $\hat{\underline{Q}}$ matrices, and compute several plausible estimates of $\underline{\theta}$, and study each. We can use these estimates as starting values for a Newton-Raphson iterative solution to the likelihood equations, or explore the likelihood function in the vicinity of these points. Wasserman (1977a) successfully analyzes two directed graphs using this strategy.

References

- Anderson, T. W. and L.A. Goodman (1957), "Statistical inference about Markov chains", Annals of Mathematical Statistics, 28, pages 89-110.
- Billingsley, P. (1961), Statistical Inferences for Markov Processes, Chicago: The University of Chicago Press.
- Bishop, Y. M. M., S. E. Fienberg, and P. W. Holland (1975), Discrete Multivariate Analysis: Theory and Practice, Cambridge, Mass.: The MIT Press.
- Darwin, J. H. (1956), "The behaviour of an estimator for a simple birth and death process", Biometrika, 43, pages 23-31.
- Elfving, G. (1937) "Zur theorie der Markoffschen ketten", Acta Social Science Fennicae u., series A.2, no. 8, pages 1-17.
- Gantmacher, F. R. (1960) The Theory of Matrices, Volume 1, New York: Chelsea.
- Holland, P. W. and S. Leinhardt (1977a), "A dynamic model for social networks" Journal of Mathematical Sociology, 5, pages 5-20.
- Holland, P. W. and S. Leinhardt (1977b), "Social structure as a network process", Zeitschrift für Soziologie, to appear.
- Keiding, N. (1974), "Estimation in the birth process," Biometrika, 61, pages 71-80.
- Keiding, N. (1975), "Maximum likelihood estimation in the birth-and-death process model", Annals of Statistics, 3, pages 363-372.
- Kendall, D. G. (1959), "Unitary dilations of one-parameter semigroups of Markov transition operators, and the corresponding integral representations for Markov processes with a countable infinity of states", Proceedings of the London Mathematical Society, 9, pages 417-431.
- Kingman, J. F. C. (1962), "The imbedding problem for finite Markov chains", Zeitschrift für Wahrscheinlichkeitstheorie, 1, pages 14-24.
- Reich, J. A. (1957), "Waiting times when queues are in tandem", Annals of Mathematical Statistics, 28, pages 768-773.

Singer, B. and S. Spilerman (1974) "Social mobility models for heterogeneous populations", Sociological Methodology, 1973-1974, edited by H. Costner. San Francisco: Jossey-Bass.

Singer, B. and S. Spilerman (1976) "The representation of social processes as Markov models", American Journal of Sociology, 82, pages 1-54.

Sylvester, J. J. (1883), "On the equation to the secular inequalities in the planetary theory", Philosophical Magazine, 16, pages 267-269.

Wasserman, S. S. (1977a), "Stochastic models for directed graphs" Ph.D. Dissertation, Department of Statistics, Harvard University.

Wasserman, S. S. (1977b), "A survey of mathematical models for graphs", Proceedings of the American Statistical Association Social Statistics Section, to appear.

Wasserman, S. S. (1977c), "Models for binary directed graphs and their applications" unpublished manuscript.